

**PENERAPAN METODE *K-MEANS CLUSTERING* DALAM  
PENGELOMPOKAN DATA PENJUALAN BARANG PADA  
PT. SWASTI TUNGGAL MANDIRI**

**Ade Febriyanti<sup>1\*</sup>, Putri Vina Bancin<sup>1</sup>, Siska Amanda<sup>1</sup>**

<sup>1</sup>Mahasiswa Prodi Sistem Informasi, STMIK Royal

\**email* : febriyantiade76@gmail.com

**Abstract:** The increasingly fierce business competition in this era of globalization requires companies to rearrange business strategies and tactics in order to always survive in business competition. To achieve this, there are several things that can be done, namely by improving product quality, adding product types, and reducing company operational costs by using company data analysis. Data mining is a statistical process that identifies and categorizes data. PT. Swasti is a company engaged in contracting and supplying the engineering needs of palm oil mills. The purpose of this research is to help classify sales data in order to maximize stock management. In this study, the application of clustering using the k-means algorithm was used. The results of the data modeling will group the stock of goods into a number of clusters. The grouping will produce information about the sale of goods where there are three groups, namely sales that are not selling well, selling well, and selling very well.

**Keywords:** k-means; data mining; clustering; sales stock.

**Abstrak:** Persaingan bisnis yang semakin ketat di era globalisasi ini menuntut perusahaan untuk menyusun kembali strategi dan taktik bisnis agar selalu bertahan dalam persaingan bisnis. Untuk mencapai hal tersebut, ada beberapa hal yang bisa dilakukan yaitu dengan meningkatkan kualitas produk, penambahan jenis produk, dan pengurangan biaya operasional perusahaan dengan cara menggunakan analisis data perusahaan. Data mining merupakan proses statistik yang mengidentifikasi dan mengelompokkan suatu data. PT. Swasti adalah perusahaan yang bergerak di bidang kontraktor dan supplier untuk kebutuhan teknik dari pabrik kelapa sawit. Tujuan diadakan penelitian ini untuk membantu mengelompokkan data penjualan agar dapat memaksimalkan manajemen stock. Pada penelitian ini, digunakan penerapan clustering dengan menggunakan algoritma k-means. Hasil dari pemodelan data tersebut akan mengelompokkan stock barang kedalam sejumlah cluster. Pengelompokkan tersebut akan menghasilkan informasi tentang penjualan barang dimana terdapat tiga cluster yaitu penjualan kurang laris, cukup laris, dan sangat laris.

**Kata kunci:** k-means; data mining; *clustering*; stok penjualan.

## PENDAHULUAN

Persaingan bisnis yang semakin ketat di era globalisasi ini menuntut perusahaan untuk menyusun kembali strategi dan taktik bisnis agar selalu bertahan dalam persaingan bisnis. Untuk mencapai hal tersebut, ada beberapa hal yang bisa dilakukan yaitu dengan meningkatkan kualitas produk, penambahan jenis produk, dan pengurangan biaya operasional perusahaan dengan cara menggunakan analisis data perusahaan[1].

Dalam kemajuan teknologi khususnya di bidang *industry*, perusahaan yang bergerak di bidang kontraktor dan supplier sangat di butuhkan. Kontraktor adalah layanan jasa konstruksi yang sangat berpengaruh dalam bidang *industry*[2]. Supplier adalah organisasi yang menyediakan sumber daya kepada customer[3].

PT. Swasti adalah perusahaan yang bergerak di bidang kontraktor dan supplier untuk kebutuhan teknik dari pabrik kelapa sawit. Dalam menjalankan usaha ini, setiap transaksi penjualan sudah menggunakan sistem komputer. Data-data penjualan tersebut tersimpan secara otomatis disebuah basis data (*database*). Dengan adanya data tersebut dapat dimanfaatkan untuk mengetahui barang yang laris di beli, cukup laris dan kurang laris di beli. Pada penelitian ini, digunakan penerapan clustering dengan menggunakan algoritma K-means. K-means merupakan algoritma pengelompokan inratif yang melakukan partisi set data ke dalam sejumlah K *cluster* yang sudah ditetapkan diawal [4].

Data mining merupakan proses statistik yang mengidentifikasi dan mengelompokkan suatu data. Data mining berguna untuk memberikan solusi kepada para pengambil keputusan dalam bisnis guna meningkatkan bisnis perusahaan[5]. Proses clustering merupakan suatu proses pengelompokan berdasarkan atas prinsip kesamaan antar kelas [6]. *Clustering* adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas atau *cluster* [7]. *Cluster* adalah kumpulan objek-objek yang sama di antara mereka dan berbeda pada objek dari cluster lainnya, K-means merupakan algoritma yang umum digunakan dalam proses *clustering* [8].

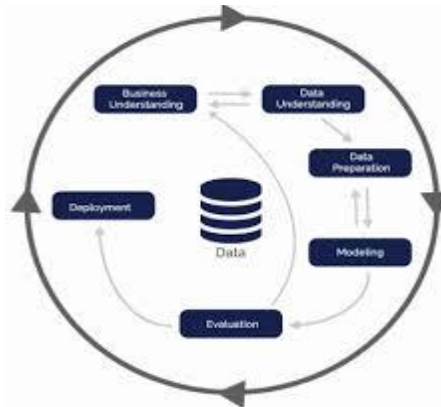
Tujuan diadakan penelitian ini untuk membantu mengelompokkan data penjualan agar dapat memaksimalkan manajemen stok. Adapun proses memaksimalkan data manajemen stok tersebut dengan cara mengambil 2 jenis atribut diantaranya yaitu *stock* awal dan keluar. Hasil dari pemodelan data tersebut akan mengelompokkan *stock* barang kedalam sejumlah *cluster*. Pengelompokkan tersebut akan menghasilkan informasi tentang penjualan barang dimana terdapat tiga *cluster* yaitu penjualan kurang laris, cukup laris, dan sangat laris.

Hasil dari pemodelan data tersebut akan mengelompokkan stock barang kedalam sejumlah cluster. Pengelompokkan tersebut akan menghasilkan informasi tentang penjualan barang dimana terdapat tiga cluster yaitu penjualan kurang laris, cukup laris, dan sangat laris.

## METODE

Metode penelitian ini dilakukan pada tahapan proses data mining yaitu *Cross-Industry Standard Process for Data Mining* (CRISP-DM)[9]. Metodologi CRISP-DM

adalah standarisasi data mining yang disusun oleh tiga penggagas. Kemudian dikembangkan pada berbagai workshop (antara tahun 1997-1999). CRISP-DM memiliki 6 tahapan analisis, yaitu *business understanding* (pemahaman bisnis), *data understanding* (pemahaman data), *data preparation* (persiapan data), *modeling* (pemodelan), *evaluation* (evaluasi), dan *deployment*(penyebaran) [10].



Gambar 1. CRISP-DM

### ***Bussines Understanding***

Dalam penelitian ini akan dilakukan analisis data penjualan di PT. Swasti agar perusahaan mengetahui data penjualan. Fokus penelitian ini adalah menganalisa bagaimana cara membuat klasifikasi untuk mengetahui stock yang ada digudang.

### ***Data Understanding***

Data sekunder yang digunakan dalam penelitian ini diperoleh langsung dari karyawan perusahaan, sebanyak 364 *record* sebagai populasi dimana data tersebut diambil pada periode bulan mei berdasarkan berkas yang ditunjukkan oleh objek penelitian. Data tersebut berisikan atribut antara lain: kode barang, deskripsi barang, stn, @, stock awal, masuk, keluar dan stock akhir.

```

: #Melihat info dataset
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 364 entries, 0 to 363
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Kode Barang           364 non-null   object
1   Deskripsi Barang      364 non-null   object
2   Stn                   364 non-null   object
3   @                     364 non-null   float64
4   Stock_Awal           364 non-null   int64
5   Masuk                17 non-null    float64
6   Keluar               335 non-null   float64
7   Stock Akhir          364 non-null   int64
dtypes: float64(3), int64(2), object(3)
memory usage: 22.9+ KB
    
```

Gambar 2. Deksripsi Data

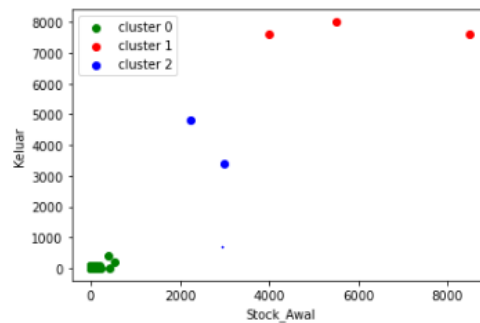
### ***Data Preparation***

Beberapa pemilihan tabel dan *field* terjadi pada fase ini. Ada beberapa proses pada tahap persiapan data antara lain, menyeleksi data, dan membersihkan data. Atribut

yang digunakan sebagai bahan data mining dibagi menjadi 2 kelompok atau 2 bagian antara lain *stock* awal, dan keluar.

**Modelling**

Dari data atribut yang telah dipilih pada fase *data preparation* yaitu *stock* awal dan keluar digunakan sebagai parameter untuk melakukan *clustering*. Hasil akhir yang akan didapatkan akan menghasilkan sebuah prediksi yang dapat dilihat dari nilai akurasi.



Gambar 3. Cluster

**Evaluation**

Pada tahap ini hasil dan validasi diukur menggunakan model clustering. *Clustering* adalah sebuah proses untuk mengelompokkan data kedalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum.

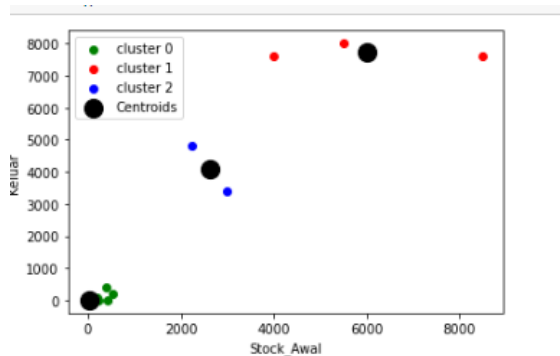
**HASIL DAN PEMBAHASAN**

Data yang digunakan dipenelitian ini merupakan data *stock* barang PT. Swasti pada bulan mei. Atribut yang dijadikan variabel dalam algoritma ini di sederhanakan menjadi atribut yang terdiri dari *stock* awal dan keluar. Tahapan awal dalam menganalisis, yaitu meng-import dataset stok gudang *palm oil mil*.

```
#Recall Dataset
df.head()
```

Kode Barang	Deskripsi Barang	Stn	@ Stock_Awal	Masuk	Keluar	Stock Akhir	
0 ABSV-0001	WATERPROOF ABRASIVE PAPER SIZE 280 X 230 MM GR...	LBR	1317.15	40	NaN	2.0	38
1 ABSV-0002	WATERPROOF ABRASIVE PAPER SIZE 280 X 230 MM GR...	LBR	703.70	20	NaN	1.0	19
2 ABSV-0004	DEPRESSED CENTER GRINDING WHEEL ULTRA TOUCH 4"...	PCS	2875.00	4	12.0	6.0	10
3 ABSV-0005	DEPRESSED CENTER GRINDING WHEEL ULTRA TOUCH 6"...	PCS	6440.00	1	12.0	1.0	12
4 ABSV-0006	DEPRESSED CENTER GRINDING WHEEL ULTRA TOUCH SIZ...	PCS	11638.88	3	12.0	5.0	10

Gambar 4. Recall Dataset



Gambar 5. Hasil Visualisasi *Cluster* dan *Centroid*

```
#melihat jumlah data dari masing-masing cluster
df['cluster'].value_counts()

Penjualan Kurang Laris    359
Penjualan Sangat Laris     3
Penjualan Cukup Laris     2
Name: cluster, dtype: int64
```

Gambar 6. Hasil *Cluster*

Berdasarkan gambar diatas, diperoleh hasil algoritma K-means dengan 3 cluster yaitu terdapat 359 penjualan kurang laris, 3 penjualan sangat laris dan 2 penjualan cukup laris.

**SIMPULAN**

Data yang digunakan dalam penelitian ini berisi 364 tabel dan 2 *field name* yang telah disortir sebelumnya yaitu *stock* awal dan keluar. Data tersebut diolah menggunakan k-means dengan bantuan *software jupyter notebook* menghasilkan item barang yang laris, cukup laris dan kurang laris sehingga pihak manajemen dapat mengelola stock barang dengan baik.

**DAFTAR PUSTAKA**

[1] M. H. Siregar, “Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan),” *J. Teknol. Dan Open Source*, vol. 1, no. 2, pp. 83–91, 2018, doi: 10.36378/jtos.v1i2.24.

- [2] D. Triana and W. O. Oktavianto, "Relevansi Kualifikasi Kontraktor Bidang Teknik Sipil Terhadap Kualitas Pekerjaan Proyek Konstruksi Di Provinsi Banten," *Fondasi J. Tek. Sipil*, vol. 2, no. 2, pp. 182–190, 2013, doi: 10.36055/jft.v2i2.1732.
- [3] H. H. Azwir, "Pemilihan Supplier Menggunakan Metode Analytic Network Process Di Pt United Tractors Pandu Engineering," *J. Tek. Ind.*, vol. 18, no. 2, p. 103, 2017, doi: 10.22219/jtiumm.vol18.no2.103-112.
- [4] D. Juliawan, F. Amir, and E. Desi, "Penerapan Data Mining Metode Clustering Pada CV. Secom Infotech Menggunakan Algoritma K-Means," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 96, 2019, doi: 10.30645/senaris.v1i0.12.
- [5] M. H. Fakhriza and K. Umam, "Analisis Produk Terlaris Menggunakan Metode K-Means Clustering Pada "Pt.Sukanda Djaya," *JIKA (Jurnal Inform.*, vol. 5, no. 1, p. 8, 2021, doi: 10.31000/jika.v5i1.3236.
- [6] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA J. Inform.*, vol. 7, no. 2, p. 109, 2019, doi: 10.30595/juita.v7i2.5529.
- [7] R. Rustam, S. Rahmatullah, S. Supriyanto, and S. Wahyuni, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Triplek Pada Pt Puncak Menara Hijau Mas," *J. Inf. dan Komput.*, vol. 8, no. 2, pp. 75–86, 2020, doi: 10.35959/jik.v8i2.186.
- [8] I. A. K. Dalam, M. Produk, T. Dan, K. Laris, P. Toko, and A. Cikarang, "Pelita teknologi," vol. 16, no. 1, pp. 79–87, 2021.
- [9] A. Prasatya, R. R. A. Siregar, and R. Arianto, "Penerapan Metode K-Means Dan C4.5 Untuk Prediksi Penderita Diabetes," *Petir*, vol. 13, no. 1, pp. 86–100, 2020, doi: 10.33322/petir.v13i1.925.
- [10] Y. P. Sari, A. Primajaya, and A. S. Y. Irawan, "Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang," *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, p. 229, 2020, doi: 10.35314/isi.v5i2.1457.